

Statistisches Protokoll / Statistical Protocol [1, 2, 4]

Das statistische Protokoll des DRRR stellt die statistischen Werkzeuge vor, die im DRRR Anwendung für die Auswertung und Beurteilungen von Datensätzen sowie Inspektionsleistungen Anwendung finden. Alle Angaben und Verweise auf diverse Normen beziehen sich immer auf die aktuell gültige Fassung.

1. Definitionen

2. Allgemeine Statistik

2.1 chemische Ringversuche

2.2 mikrobiologische Ringversuche

2.3 sensorische Ringversuche

2.3.1 Rangordnungsprüfung

2.3.2 Dreiecksprüfung

2.3.3 Wasserbeschaffenheit

2.3.4 Einfach beschreibende Prüfung

2.3.5 Sensorik von Packstoffen und Packmitteln, Sensorik von Papier und Pappe

2.4 kleine Ringversuche „small ILC“

3. Homogenität und Stabilität

4. Messung der Laborbewertung

5. Umgang mit Kundendaten

The statistical protocol presents the tools applied at DRRR for evaluating datasets and inspection purposes. Every information and all references to various standards / norms refer to the currently valid version.

1. Definitions

2. General statistics

2.1 Chemical Proficiency Tests

2.2 Microbiological Proficiency Tests

2.3 Sensory Proficiency Tests

2.3.1 Ranking test

2.3.2 Triangle test

2.3.3 Water quality

2.3.4 Simple descriptive testing:

2.3.5 Sensory of packaging and sensory of paper and board

2.4 “Small ILC’s”

3. Homogeneity and stability

4. Measurements of performance

5. Handling of customer data

1. Definitionen

1. Definitions

Begriff	Erklärung	Abk./ abbr.	term	explanation
Einzelwert	Einzelwert einer Mehrfachbestimmung	x	single value	single value of a multi determination
Labormittelwert	Mittelwert einer Mehrfachbestimmung	m	lab mean value	mean of a multi determination
Mittelwert Datensatz		m_{all}	mean of a data set	
Bester Schätzwert für den wahren Wert bei Ringversuchen (RV)	Für die Berechnung des Besten Schätzwertes für den wahren Wert werden die Ergebnisse der Teilnehmer herangezogen, die die Referenzmethode (Bezugsmethode(n)) benutzt haben. Der m _{best} wird somit auf die Bezugsmethode(n) metrologisch rückgeführt. Die Berechnung erfolgt entsprechend der ausgewählten Statistik.	m_{best}	best estimate of the true value in proficiency testing schemes (RV)	For calculating the best estimate of the true value only results of the participants using the reference (DRRR- standard method(s)) method are considered. So m _{best} is traced back metrologically to the DRRR standard method(s). The calculation is computed according the selected statistical procedure.
Referenzwert	Bei einem Referenzmaterial RM ist der zugewiesene Wert der Referenzwert	m_{ref}	reference value	For reference material RM the assigned value is the reference value
Vergleichsstandardabweichung	Ist in der jeweiligen Norm/Standard festgelegt, oder wird gemäß ISO 5725-2 (Gleichung 24) berechnet	s_R	reproducibility standard deviation	Defined in each norm/standard or calculated according ISO 5725-2 (equation 24)
Wiederholstandardabweichung	Ist in der jeweiligen Norm/Standard festgelegt, oder wird gemäß ISO 5725-2 (Gleichung 20) berechnet	s_r	repeatability standard deviation	Defined in each norm/standard or calculated according ISO 5725-2 (equation 20)
Wiederholbarkeit	Ist in der jeweiligen Norm/Standard festgelegt, oder wird aus s_r durch Multiplikation mit 2,8 berechnet	r	repeatability	Defined in each norm/standard or calculated out of s_r by multiplying with 2,8
Relative Wiederholbarkeit	Die relative Wiederholbarkeit bezieht sich auf m_{best} und wird in % ausgedrückt	$r_{rel}(\%)$	Relative repeatability	The relative repeatability is expressed in % with respect to m_{best}
Vergleichbarkeit	Ist in der jeweiligen Norm/Standard festgelegt, oder wird aus s_R durch Multiplikation mit 2,8 berechnet	R	reproducibility	Defined in each norm/standard or calculated out of s_R by multiplying with 2,8
Relative Vergleichbarkeit	Die relative Vergleichbarkeit bezieht sich auf m_{best} und wird in % ausgedrückt	$R_{rel}(\%)$	Relative reproducibility	The relative reproducibility is expressed in % with respect to m_{best}

Standard-abweichung	Ist ein Maß für die Streuung um einen Mittelwert ausgedrückt	s	standard deviation	Expresses the deviation around the mean
Standard-abweichung der analytischen Homogenität	Ist ein Maß für die Streuung von 10 Untersuchungen an 1 repräsentativen Probe um den zugehörigen Mittelwert ausgedrückt	S_{analyt}	standard deviation of the analytical homogeneity	Expresses the deviation around the mean of 10 testing results of 1 representative sample around the corresponding mean
Standard-abweichung der Stichprobenhomogenität	Ist ein Maß für die Streuung von 10 repräsentativen Stichproben um den zugehörigen Mittelwert ausgedrückt	S_{Stichprobe}	standard deviation of the random sample homogeneity	Expresses the deviation around the mean of 10 representative random samples around the mean
Standard-abweichung der Materialhomogenität	Ist ein Maß für die Streuung der Ergebnisse resultierend aus der Berechnung von S_{analyt} und $S_{Stichprobe}$	S_{Material}	standard deviation of the material homogeneity	Is an expression for the deviation of the results derived from the calculation of S_{analyt} and $S_{Stichprobe}$
Standard-abweichung RV / mittlere absolute Abweichung RV (abhängig von angewandter Statistik)	Ist ein Maß für die Streuung der Ergebnisse um den Besten Schätzwert für den wahren Wert	S_{best}	standard deviation RV / mean average deviation RV (dependent on the applied statistics)	Expresses the deviation around the best estimate for the true value
(Maximale) Standardabweichung zur Eignungsbeurteilung	Ist ein Maß für die maximale Streuung der Ergebnisse um den Besten Schätzwert für den wahren Wert	SDPA	maximum standard deviation for proficiency assessment	Expresses the maximum deviation around the best estimate for the true value
Minimale Standardabweichung zur Eignungsbeurteilung	Ist ein Maß für die minimal Streuung der Ergebnisse um den Besten Schätzwert für den wahren Wert	SDPA_{min}	minimum standard deviation for proficiency assessment	Expresses the minimum deviation around the best estimate for the true value
Standard-abweichung RM	Ist ein Maß für die Streuung der Ergebnisse um den zugehörigen Referenzwert	S_{ref}	standard deviation RM	Expresses the deviation around the reference value
Erweiterungsfaktor	Erweiterungsfaktor zur Bestimmung der erweiterten Unsicherheit (95,5%)	k=2	coverage factor	Coverage factor for calculating the extended uncertainty (95,5%)
Studentfaktor	Erweiterungsfaktor für kleine Datensätze (abhängig von der Größe des Datensatzes), Anwendung wie <i>k</i> .	t	student factor	Coverage factor for calculating the extended uncertainty for small data sets (<i>t</i> depends on size of data set) - applied similar to <i>k</i> .
erweiterte Unsicherheit (95,5%)	erweitertes Unsicherheitsintervall (95,5% Wahrscheinlichkeit) um einen Mittelwert ausgedrückt in Form der Standardabweichung multipliziert mit dem entsprechenden <i>k</i> bzw <i>t</i>	U_{ex}	extended uncertainty	extended uncertainty interval (95,5% probability) around the mean in terms of the standard deviation multiplied by the corresponding <i>k</i> or <i>t</i> .
Unsicherheit (95,5%)	Unsicherheitsintervall (95,5% Wahrscheinlichkeit) um den zugehörigen Mittelwert ausgedrückt	u	uncertainty (95,5%)	Uncertainty (95,5% probability) around the mean expressed as a confidence interval (95,5%)

	drückt in Form des Vertrauensbereichs (95,5%)			around the related mean
Unsicherheit (95,5%) des besten Schätzwertes	Unsicherheitsintervall (95,5% Wahrscheinlichkeit) um den zugehörigen Besten Schätzwert (m_{best}) ausgedrückt in Form des Vertrauensbereichs (95,5%)	U_{best}	uncertainty (95,5%) of the best estimate	Uncertainty interval (95,5% probability) around the best estimate (m_{best}) expressed as a confidence interval (95,5%)
Unsicherheit (95,5%) des Referenzwertes	Unsicherheitsintervall (95,5% Wahrscheinlichkeit) um den zugehörigen Referenzwert (m_{ref}) ausgedrückt in Form des Vertrauensbereichs (95,5%)	U_{ref}	uncertainty (95,5%) of the reference value	Uncertainty interval (95,5% probability) around the reference value (m_{ref}) expressed as confidence interval (95,5%)
Stichproben-Homogenität	Stichproben-Homogenität ist das Intervall eines erwarteten Messbereichs ausgedrückt als erweiterte Unsicherheit (95,5%) U_{ex} von jeweils einer Einzeluntersuchungen von 10 repräsentativen Stichproben	H_s	homogeneity of random sample batch	This total uncertainty (H_s) expressed as extended uncertainty (95,5%) U_{ex} of 10 measurements out of 10 representative random samples (in between homogeneity).
Analytische Homogenität	Analytische Homogenität der Einzelprobe ist das Intervall eines erwarteten Messbereichs ausgedrückt als erweiterte Unsicherheit (95,5%) U_{ex} von 10 Einzeluntersuchungen an 1 repräsentativen Stichprobe. Da angenommen werden kann, daß eine Einzelprobe ideal homogen ist, resultiert die tatsächlich gemessene erweiterte Unsicherheit (95,5%) U_{ex} ausschließlich aus der analytischen Messunsicherheit.	H_a	analytical homogeneity	The analytical uncertainty (H_a) expressed as extended uncertainty (95,5%) U_{ex} of 10 measurements out of 1 representative random sample. It is assumed that one sample itself shows no inhomogeneity.
Material Homogenität	Material-Homogenität ist das resultierende erweiterte Unsicherheitsintervall (95,5%) U_{ex} aus der Differenz der Wurzel der quadrierten Standardabweichungen s ermittelt aus der Bestimmung der Stichproben-Homogenität (H_s) und der Analytischen Homogenität (H_a).	H_M	material homogeneity	The material homogeneity expresses the extended uncertainty (95,5%) U_{ex} . It is computed out of the square root of the squared homogeneity of the random sample batch (H_s) and analytical homogeneity (H_a).
Anzahl der Elemente eines Datensatzes	in der Sensorik Anzahl an verkosteten Triaden	n	number of elements of a data set	In sensory evaluation number of tasted triads
CRD-Wert	kritischer Differenzwert	CRD	CRD value	Critical difference value
CD	kritische Differenz (wird aus s_R und s_r berechnet)	CD	CD	Critical difference (calculated by s_R and s_r)
Median	Zentralwert	med	Median	Central value
Mittlere absolute		MAD	Median absolute	

Abweichung			deviation	
Normalisierter Quartilsabstand		nIQR	Normalised inter-quartile range	
Arithmetisches Mittel	Durchschnittswert		Arithmetic mean	Average
Wahrscheinlichkeit für das Eintreten eines Zustands		P	Probability for the occurrence of state	
Erfolgswahrscheinlichkeit	z.B. In der sensorischen Dreiecksprüfung beträgt die Erfolgswahrscheinlichkeit 1/3	p	probability for success	e.g. in the sensory triangle test the probability for success is 1/3
Wahrscheinlichkeit für einen Misserfolg	z.B. In der sensorischen Dreiecksprüfung beträgt die Wahrscheinlichkeit für einen Misserfolg 2/3	q	probability for failure	e.g. in the sensory triangle test the probability for success is 2/3
Anzahl der Erfolge	z.B. In der sensorischen Dreiecksprüfung die Anzahl an richtig bewerteten Triaden	y	Number of successes	e.g. in the sensory triangle test the number of correctly judge triads

2. Allgemeine Statistik

2. General statistics

Sensible Statistik / sensible statistics

Mittelwert

mean

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

Standardabweichung

standard deviation

$$s = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n (m - m_{all})^2}$$

Unsicherheit (95,5 %)

uncertainty (95,5 %)

$$u = \frac{s \cdot t}{\sqrt{n}}$$

Ausreißertest nach Grubbs [4]

outlier test (Grubbs) [4]

Die Berechnung des Grubbs-Ausreißertests erfolgt in mehreren Schritten:

Schritt 1) Berechnung der Differenz von **Labormittelwert – Gesamtmittelwert = d_i** (des i-ten Labors)

Schritt 2) Berechnung der **Puffergröße = d_i / Standardabweichung**

Schritt 3) Vergleich der Puffergröße mit Grubbs-Werten (Abhängig von Anzahl der Elemente eines Datensatzes **n**)

Schritt 4) Ein **Ausreißer** liegt vor, wenn die **Puffergröße größer ist als der Grubbs-Wert**

Kein Ausreißer liegt vor, wenn die Puffergröße kleiner ist als der Grubbs-Wert.

The calculation of the Grubbs outlier test happens in different steps:

Step 1) Calculation difference of **laboratory mean – mean total = d_i**

Step 2) Calculation of **buffer size = d_i / standard deviation**

Step 3) By comparing buffer size with Grubbs-values (dependent on number of elements of a data set **n**)

Step 4) An **outlier exists**, if the **buffer size is bigger than the Grubbs-value**.

The results are acceptable, if the buffer size is smaller than the Grubbs-value

Ausreißertest nach Bremser

Die Berechnung des Bremser-Ausreißertests erfolgt in mehreren Schritten:

Schritt 1) Berechnung der Varianz der beiden Mittelwerte der Doppelbestimmung (s^2_{bwi}) (Stichproben-Streuung des Labors)

Schritt 2) Berechnung der laborspezifischen Messunsicherheit aus den beiden Doppelbestimmungen eines Labors

$$= \sqrt{(s_{D1}^2 + s_{D2}^2) / 2}$$

= analytische Streuung des Labors)

Schritt 3) Berechnung Varianz unter Berücksichtigung der Materialstreuung, analytische Streuung des Labors und der F-Verteilung (da Vergleich verschiedener Streuungen)

$$F_{0.05,1,48} \cdot (s_{in\ hom}^2 + s_{repeat,i}^2)$$

Schritt 4) Kein Ausreißer liegt vor, wenn Stichproben-Streuung des Labors kleiner ist

$$s_{bw,i}^2 \leq F_{0.05,1,48} \cdot (s_{in\ hom}^2 + s_{repeat,i}^2)$$

Outlier test (Bremser)

The calculation of the Bremser outlier test happens in different steps:

Step 1) Calculation of the variance of both mean values of a double determination (s^2_{bwi}) (batch spreading of the laboratory)

Step 2) Calculation of laboratory-specific measurement uncertainty of the double determination of a single laboratory

$$= \sqrt{(s_{D1}^2 + s_{D2}^2) / 2}$$

= analytical spreading of the laboratory)

Step 3) Calculation of the variance considering the material spreading, analytical spreading of the laboratory and the F-distribution (because comparison of different spreadings)

$$F_{0.05,1,48} \cdot (s_{in\ hom}^2 + s_{repeat,i}^2)$$

Step 4) No **outlier exists**, if the s^2_{bwi} is smaller

$$s_{bw,i}^2 \leq F_{0.05,1,48} \cdot (s_{in\ hom}^2 + s_{repeat,i}^2)$$

Offensichtlich grobe Fehler „Blunder“

Offensichtlich grobe Fehler / „Blunder“ sollten keiner Ausreißerprüfung (Grubbs) oder robusten Methoden ausgesetzt werden. Deshalb wird vor der Anwendung der statistischen Methoden der Datensatz auf mögliche offensichtlich grobe Fehler / „Blunder“ geprüft.

Wir unterscheiden zwischen:

a.) technischer Ausreißer

Tippfehler, falsche Einheit, Dezimalstellenfehler, Proben wurden vertauscht, etc.

b.) Expertenausreißer

Zwei Bewertungskriterien werden zur Identifizierung angewendet, beide müssen erfüllt sein:

1. Labormittelwert liegt außerhalb von Median \pm 50%
2. Labormittelwert liegt mehr als die 5-fache robuste Standardabweichung. vom robusten Mittelwert weg.

Obviously gross errors “Blunder”

Obviously gross errors / “blunders” should not be subjected to outlier testing (Grubbs) or robust methods. Therefore before applying the statistical methods, the data set is checked for possible obviously gross error / “blunders”.

We differentiate between:

a.) technical outlier

Typing errors, incorrect unit, decimal point error, samples are mixed-up and so on

b.) Expert outlier

Two evaluation criteria are used to identify an expert outlier, both must be met:

1. The laboratory mean lies outside the median \pm 50%
2. The laboratory mean value is more than 5 x robust standard deviation away from the robust mean value.

Ausreißertest s_r – Ausreißer	Outlier test s_r – outlier
<p>Die Berechnung des s_r – Ausreißertests erfolgt in mehreren Schritten:</p> <p>Schritt 1) Berechnung von s_r gemäß Formeln für Präzisionsdaten oder Entnahme aus der Bezugsmethode wenn vorhanden.</p> <p>Schritt 2) Vergleich von s_r mit der Standardabweichung der einzelnen Labore.</p> <p>Schritt 3) Ein Ausreißer liegt vor, wenn die Laborstandardabweichung s_L größer ist als der zweifache s_r-Wert.</p> <p>Kein Ausreißer liegt vor, wenn die Laborstandardabweichung kleiner ist als der 2-fache s_r-Wert.</p> <p>Der Ausreißertest wird nur einmalig angewendet. Eventuell neu auftretende Ausreißer nach Ausreißer Eliminierung der ersten Anwendung werden nicht berücksichtigt. Eine Neuberechnung der Gubbstest erfolgt nach der Neuberechnung des Lage- und Streumaßes nicht.</p>	<p>The calculation of the s_r – <i>outliertest</i> happens in different steps:</p> <p>Step 1) Calculation of s_r according to the formula for precision data or taken from the standard method if available.</p> <p>Step 2) By comparing s_r with the standard deviation of the single laboratories.</p> <p>Step 3) An outlier exists, if the lab standard deviation s_L is bigger than the double s_r-value.</p> <p>The results are acceptable, if the lab standard deviation s_L is smaller than the double s_r-value</p> <p>The outlier test is done only once. Additional outliers after outlier elimination of the first procedure are not considered. An additional calculation of the Grubbs test after the new calculation is not done.</p>

Weitere Expertenausreißer	Further expert outliers
<p>Es gibt die Möglichkeit weitere Expertenausreißer auf Grundlage einer Expertenentscheidung festzulegen. Expertenausreißer sind z.B. Ausreißer, die nicht vom Grubbs-Ausreißertest als Ausreißer erkannt werden, weil sie durch Ihre Lage innerhalb des Datensatzes den Grubbs-Ausreißertest zum Scheitern bringen. Dies kann beispielsweise dann vorkommen, wenn zwei Werte nahe beieinander liegen. Die zwei einzelnen Werte alleine würden vom Grubbs-Ausreißertest jeweils als Ausreißer erkannt werden, beide Werte zusammen, legen den Grubbs-Ausreißertest lahm.</p>	<p>It is possible to define further expert outliers on the basis of expert decisions. Expert outliers are for example outliers that are not recognized as outliers by the Grubbs outlier test, because their position within the data causes the Grubbs outlier test to fail. This can occur, for example, if two values are close together. The two individual values alone would be recognized as outliers by the Grubbs outlier test, both values together would paralyze the Grubbs outlier test.</p>

Robuste Statistik / robust statistics

Median	median
$\text{med} = \left\{ \frac{1}{2} \left(\text{med} \left(\frac{n}{2} \right) + \text{med} \left(\frac{n}{2} + 1 \right) \right) \right\}$	$\text{wennAnzahlgerade} = \{x x = 2n \wedge n \in \mathbb{N}\}$
$\text{med} = \left\{ \text{med} \left(\frac{n+1}{2} \right) \right\}$	$\text{wennAnzahlungerade} = \{x x = 2n + 1 \wedge n \in \mathbb{N}\}$

Mittlere absolute Abweichung vom Median	median absolute deviation
<p>Der MAD ist ein robuster Schätzer für die absolute Abweichung vom Median. Sie berechnet sich aus dem inneren Median $med_i(x_j)$, also dem Median von n Werten und dem äußeren Median, dem Median der Differenzen vom inneren Median und den einzelnen Werten multipliziert mit dem Faktor 1,4826</p>	<p>A very robust scale estimator is the MAD, the median absolute deviation from the median (Hampel, 1974), which is computed as where the inner median, $med_i(x_j)$, is the median of the n observations, and the outer median (taken over i) is the median of the n absolute values of the deviations about the inner median multiplied with factor 1,4826</p>
$MAD = 1,4826 \text{ med}_i \left(x_i - \text{med}_j(x_j) \right)$	

Normalisierter Quartilabstand (nIQR)	Normalized interquartile range (nIQR)
<p>Sind mindestens 50% der Messwerte der Teilnehmer gleich wird der MAD zu Null. In diesem Fall wird anstelle des MAD der normalisierte Quartilabstand (nIQR) berechnet.</p>	<p>If 50% or more of the participant results are the same, then MAD will be zero. In this case the Normalized interquartile range (nIQR) is calculated instead of the MAD.</p>
$nIQR(x) = 0,7413 (Q_3(x) - Q_1(x))$	

Robuster Mittelwert ^[4, 5]	robust mean ^[4, 5]
Hampelschätzer	Hampel estimator
$\sum_{j=1}^J \Psi \left(\frac{y_j - m}{s_R} \right) = 0 \quad \Psi(x) = \left\{ \begin{array}{ll} 0 & x \leq -4,5 \\ -4,5 - x & -4,5 < x \leq -3 \\ -1,5 & -3 < x \leq -1,5 \\ x & -1,5 < x \leq 1,5 \\ 1,5 & 1,5 < x \leq 3 \\ 4,5 - x & 3 < x \leq 4,5 \\ 0 & x > 4,5 \end{array} \right\}$	

Robuste Standardabweichung [4, 5]

Q-Methode

robust standard deviation [4, 5]

Q-method

$$s_{robust} = \frac{G_1^{-1}(q)}{\sqrt{2\Phi^{-1}(0,5 + 0,5q)}}$$

Berechnung von Präzisionsdaten / Calculation of precision data

Bei Bedarf werden Präzisionsdaten berechnet, z.B. wenn in der jeweiligen Bezugsmethode keine oder nicht anwendbare Daten hinterlegt sind. Für Auswertungen im Bereich der Materialprüfung werden wann immer möglich Präzisionsdaten berechnet.

If necessary, precision data are calculated, for example, if no data or data that cannot be used is mentioned in the relevant reference method. For evaluations in the field of material testing, precision data is calculated whenever possible.

Gesamtmittelwert

standard deviation

$$\bar{y}_j = \frac{\sum_{i=1}^p n_{ij} \bar{y}_{ij}}{\sum_{i=1}^p n_{ij}}$$

Wiederholvarianz s_r^2

repeatability variance s_r^2

$$s_{rj}^2 = \frac{\sum_{i=1}^p (n_{ij} - 1) s_{ij}^2}{\sum_{i=1}^p (n_{ij} - 1)}$$

Varianz zwischen den Labors s_L^2

between-laboratory variance s_L^2

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{n_j}$$

wobei / where

$$s_{dj}^2 = \frac{1}{p-1} \sum_{i=1}^p n_{ij} (\bar{y}_{ij} - \bar{y}_j)^2$$

und / and

$$n_j = \frac{1}{p-1} \left[\sum_{i=1}^p n_{ij} - \frac{\sum_{i=1}^p n_{ij}^2}{\sum_{i=1}^p n_{ij}} \right]$$

Vergleichsvarianz s_R^2

reproducibility variance s_R^2

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2$$

Wiederholgrenze r

repeatability limit r

$$2,8 * s_r$$

Vergleichsgrenze R

reproducibility limit R

$$2,8 * s_R$$

2.1 Verwendete Statistik in der Ringversuchsauswertung (CHEMIE)

2.1 applied statistics for evaluation of proficiency testing schemes (CHEMISTRY)

Ringversuche werten wir prinzipiell mit drei verschiedenen statistischen Modellen aus. Der Grund dafür ist, dass es nicht die ideale Statistik gibt, genauso wie es den idealen Datensatz für eine Auswertung nicht gibt.

Ebenso wenden wir vorzugsweise den Consensus-Value zur Bestimmung des mbest an und nicht den Wert eines einzelnen Referenzlabors. Dies liegt darin begründet, dass wir nicht ein einziges „Referenzlabor“ über andere Labore stellen sondern immer eine Gruppe von Laboratorien, die am jeweiligen Ringversuch teilgenommen haben für die Festlegung von mbest bevorzugen. Somit ist der Bezug zur Probe gesichert und der Fehler eines einzelnen Labors wirkt sich nicht wesentlich aus.

Die folgenden 3 Verfahren finden immer Anwendung in der Auswertung:

Sensible Statistik

Sensible Statistik mit Ausreißereliminierung (nach Grubbs, ISO 5725)

Robuste Statistik

Die Auswertung der Daten kann methodenabhängig und/oder methodenunabhängig erfolgen, abhängig von der Art des jeweiligen Ringversuchs.

Besonderheiten bei Ringversuchen zur Umweltsimulation (z.B. Lichtechtheit)

Nach Durchführung der Bewitterungsprüfung und Auswertung der Proben durch die Labore (z.B. Graumaßstab oder Farbmessung) werden diese zum DRRR geschickt und können so noch einmal „zentralisiert“ ausgewertet werden. Hierzu werden die geprüften Proben zu einem Expertenlabor (Auswahl wurde im Vorfeld festgelegt) geschickt und dort z.B. auf die Farbänderung untersucht. Aus den Messwerten des Expertenlabors wird dann wiederum der Consensus-Value berechnet, das Expertenlabor gibt keine Mittelwertlage vor.

Interlaboratory comparisons are evaluated by three different statistical schemes because it do not exist the ideal statistical scheme as well as the ideal dataset for an evaluation.

We apply the consensus value for determination of the assigned value (mbest) and not the result of a single “reference lab”. We do so because we do not see a single “reference lab” above others but prefer many labs. Following the possible error of one lab does not influence mbest significantly

The following three statistical methods are always used for evaluation:

sensible statistics

sensible statistics with outlier elimination (acc. to Grubbs, ISO 5725)

robust statistics

The evaluation of the data can be method-dependent and/or method-independent, depending on the type of the respective proficiency test.

Special features of proficiency tests for environmental simulation (e.g. light fastness):

After the weathering test has been performed and the samples have been evaluated by the laboratories (f.e. grey scale or colour measurement), they are sent to the DRRR and can thus be evaluated again in a "centralized" manner. For this purpose, the tested samples are sent to an expert laboratory (see valid white Exce for selection) where they are examined, for example for colour changes. The consensus value is then calculated from the measured values of the expert laboratory; the expert laboratory does not specify a mean value.

Chi-Quadrat-Anpassungstest

Mit Hilfe des χ^2 -Anpassungstests werden Datensätze auf die Normalverteilung geprüft. Dazu werden die Datensätze in 6 symmetrische Klassen eingeteilt. Für die Klassenelemente werden Erwartungswerte berechnet und absolute Häufigkeiten bestimmt. Daraus errechnet sich ein χ^2 -Wert, mit dem die Nullhypothese auf Normalverteilung angenommen (Datensatz ist normalverteilt) bzw. abgelehnt (Datensatz ist nicht normalverteilt) werden kann. Mit den berechneten χ^2 -Werten lässt sich darüber hinaus der Grad der Normalverteilung eines Datensatzes bestimmen.

chi-squared goodness of fit test

The chi-squared goodness of fit test is used to test if datasets follow a normal distribution. The test procedure consists of generation of classes, generation of expected values and observed frequencies.

Accordingly a chi-squared value is calculated, which is used to accept (dataset follows normal distribution) or reject (dataset does not follow a normal distribution) the null hypothesis. Additionally with the χ^2 -value one can determine the degree of normal distribution of a dataset.

Qualitätsmessung und Auswahl der drei statistischen Verfahren

Zur Überprüfung der Qualität der statistischen Verfahren wird der χ^2 -Anpassungstest durchgeführt. Dieser Test bestimmt über den jeweiligen Mittelwert und Standardabweichung der unterschiedlichen statistischen Verfahren die dazugehörigen Dichtefunktionen. Diese Funktionen werden mit idealen Dichtefunktionen verglichen. Je größer die Abweichung von idealer zu realer Dichtefunktion ist, umso geringer ist der normalverteilte Anteil von Daten in einem Datensatz. Die Abweichung wird mit dem χ^2 -Wert bestimmt. Vergleicht man die verschiedenen χ^2 -Werte der drei statistischen Verfahren, kann man beurteilen, welches der Verfahren den normalverteilten Anteil im Datensatz am besten erkannt hat. Für unseren Einsatzzweck wird über dieses Verfahren die Qualität der drei Statistiken beurteilt. Je kleiner der χ^2 -Wert, desto „besser“. Ist der χ^2 -Wert größer als 7,82 so ist der Datensatz als nicht normalverteilt erkannt worden und das statistische Verfahren ist für den entsprechenden Datensatz nicht sinnvoll anwendbar. Ist keines der drei statistischen Verfahren geeignet, wird der Median zur Bestimmung des besten Schätzwertes verwendet.

Ausnahme 1: Ringversuchsauswertung bei diskreten / nicht kontinuierlichen Messwerten

Liegen diskrete / nicht kontinuierlichen Messwerte für einen Parameter vor ist die Annahme einer Normalverteilung nicht zielführend. Aus diesem Grund wird für alle Parameter mit diesen Messwerten unabhängig vom χ^2 -Wert der Median (med) zur Bestimmung

quality measurement and selection of the three statistical methods

The chi-square-goodness of fit test is used to proof the quality of the evaluation's statistical scheme. This test defines the density function of all three statistical methods with the associated mean and standard deviation. These density functions get compared with an ideal density function. The more the real density function differs from the ideal function, the lower is the chance of a dataset's normal distribution.

The difference is defined by a chi-square-value. To get information about the statistical scheme which identified the normal distribution the best way, the three different chi-square values get compared. This way assesses the three statistical method's quality. The lower the chi-square-value, the "better" the normal distribution of the dataset. If this value is greater than 7,82 the dataset will not recognized as normal distributed and the associated statistical method is not representative for the corresponding dataset. If the different used statistical methods fail, then the median will be used to define the best estimate.

Exception 1: evaluation of proficiency testing with discrete / non continuous data sets

If a data set has discrete / non continuous measurement values the assumption of a normal distribution is not expedient. For those data the statistical evaluation is done, regardless of the χ^2 -value, with the median (med) to define the best estimate. In this

des besten Schätzwertes verwendet. In diesem Fall ist die mittlere absolute Abweichung (MAD) ein Maß für die Streuung der Messwerte.

Sind mindestens 50% der Messwerte der Teilnehmer gleich wird der MAD zu Null. In diesem Fall wird anstelle des MAD der normalisierte Quartilabstand (nIQR) berechnet.

Wird auch der nIQR zu Null kann eine alternative Berechnung, entweder durch robuste Standardabweichung mittels Q-Methode, arithmetische Standardabweichung (nach Ausreißereliminierung) oder arithmetische Standardabweichung (ohne Ausreißereliminierung) als Maß für die Streuung der Messwerte genutzt werden. Beispiele für diskrete / nicht kontinuierliche Messwerte:

- Graumaßstab (z.B. Farbänderung, Anbluten)
- Blaumaßstab (z.B. Farbänderung)
- Notensysteme (z.B. Geruch)
- Klassen (z.B. Korrosion, Rostgrad)
- Vergleichsbilder (z.B. Blasengrad)

Ausnahme 2: Anzahl Ausreißer größer als 25%

Hat die sensible Statistik ohne Ausreißer den geringsten χ^2 -Wert und es wurden mehr als 25% der Labore als Ausreißer (z.B. Grubbs oder sr-Ausreißer) erkannt, sollte diese Statistik nicht für den Ringversuch herangezogen werden. Es sollte die Statistik mit dem nächst höheren χ^2 -Wert ausgewählt werden.

Ausnahme 3: Berücksichtigung der relativen Standardabweichung

Hat die Statistik mit dem geringsten χ^2 -Wert eine relative Standardabweichung, die größer als 50% vom besten Schätzwert ist wird diese Statistik nicht für den Ringversuch herangezogen. Es wird die Statistik mit dem nächst höheren χ^2 -Wert ausgewählt. Haben alle zur Verfügung stehenden Statistiken eine relative Standardabweichung von $\geq 50\%$ des besten Schätzwertes wird für diesen Parameter keine Laborbewertung stattfinden. In diesem Fall werden die Messdaten zwar ausgewertet und dargestellt, aber es wird keine Laborbewertung durchgeführt.

Bei Datensätzen mit diskreten / nicht kontinuierlichen gilt diese Regelung nicht, da keine Normalverteilung angenommen werden kann.

case the median absolute Deviation (MAD) provides an estimate of the population standard deviation.

If 50 % or more of the participant results are the same then MAD will be zero. In this case the Normalized interquartile range (nIQR) is calculated instead of the MAD.

If the nIQR will be zero as well an alternative calculation for the population standard deviation is done by using robust standard deviation (Q-method), arithmetic standard deviation (after outlier removal) or arithmetic standard deviation (without outlier removal). Example discrete / non continuous measurement values:

- grey scale (e.g. change in color, staining)
- blue scale (e.g. change in color)
- notes (e.g. odour)
- classes (e.g. corrosion, degree of corrosion)
- reference pictures (e.g. Blistering)

Exception 2: Number of outliers greater than 25%

If the sensitive without outliers has the lowest χ^2 value and more than 25% of the laboratories were identified as outliers (e.g. Grubbs or sr outliers), this statistic shall not be used for the proficiency testing. The statistic with the next higher χ^2 value shall be selected.

Exception 3: Consideration of the relative standard deviation

If the statistic with the lowest χ^2 value has a relative standard deviation greater than 50% of the best estimate, this statistic will not be used for the PT. The statistic with the next higher χ^2 value will be selected. If all available statistics have a relative standard deviation of $\geq 50\%$ of the best estimated value, no laboratory evaluation may take place for this parameter. The evaluation for this parameter is displayed informatively in the report, z-score or z'score evaluations are not allowed. In this case the dataset is evaluated and the results are presented, but an evaluation of laboratory performance is not done.

For data sets with discrete measurement values this regulation is not valid, due to not having normal distributed data.

**Verwendete Statistik in der
Ringversuchsauswertung
(CHEMIE-QUALITATIV)**

**Applied statistics for evaluation of
proficiency testing schemes
(CHEMISTRY-QUALITATIVE)**

Ziel der qualitativen Prüfung ist es, dass alle 3 qualitativen Proben richtig bestimmt werden (positiv oder negativ). Werden nur 2 von 3 Proben richtig bestimmt gilt die Aufgabenstellung bereits als nicht erfolgreich gelöst.

Die Probenstruktur wurde so gewählt, dass die Wahrscheinlichkeit ein vollständig richtiges Ergebnis zu raten relativ unwahrscheinlich ist (12,5%). Die Wahrscheinlichkeit alle Ergebnisse richtig zu raten berechnet sich nach der Bernoulli-Gleichung.

Ringversuche zur Identifizierung (z.B. Monofilien, Mehrschichtfolien, Kunststoffgranulat, Textilfasern):

Ungleich einer qualitativen Fragestellung (positiv oder negativ) handelt es sich bei den Ringversuchen zur Identifikation nicht nur um ein positives oder negatives Ergebnis eines vorab bekannten Materials, sondern jede Probe muss ohne Vorkenntnisse über die möglichen eingesetzten Materialien korrekt identifiziert werden. Es werden unterschiedlich viele verschiedene Proben im Ringversuch eingesetzt (üblich 3-6). Eine Identifizierung stellt also komplexere Ansprüche an die Untersuchung und das Endergebnis, sodass an dieser Stelle nicht die Bernoulli-Gleichung wie im Falle von positiven/ negativen Ergebnissen zur Auswertung herangezogen werden kann.

Ziel der Prüfung ist es immer alle zur Verfügung gestellten Proben korrekt zu identifizieren. Wird nur eine der Proben nicht korrekt identifiziert, so gilt die Aufgabenstellung als nicht erfolgreich gelöst.

The aim of this qualitative test is to ensure that all qualitative samples are correctly determined (positive or negative). If only 2 out of 3 samples are determined correctly, the task is already considered to have been unsuccessfully solved.

The sample structure was chosen in a way, that the probability of guessing a completely correct result is relatively unlikely (12.5%). The probability to guess a results correctly is calculated according to the Bernoulli equation.

Proficiency testing for identification (e.g. monofilms, multilayer films, plastic pellets, textile fibres):

Unlike of the qualitative question (positive or negative) the proficiency tests for identification are not only a positive or negative result of a previously known material, but each sample must be correctly identified without prior knowledge of the possible materials used. A different number samples are used in the proficiency test (normally 3-6). An identification therefore make more complex demands on the examination and the final result, so that the Bernoulli equation cannot be used here for evaluation as in the case of positive/negative results.

The aim of this test is to correctly identify all available samples with regard to the material used. If for example only 2 of the 3 samples are correctly identified, the task is considered to have been unsuccessfully solved

2.2 Verwendete Statistik in der Ringversuchsauswertung (MIKROBIOLOGIE)

2.2 applied statistics for evaluation of proficiency testing schemes (MICROBIOLOGY)

Bei Anwendung des Bremser-Ausreißers als statistisches Werkzeug und damit einhergehender Probenstruktur wird nur die sensible Statistik (mit Ausreißer-Eliminierung) angewandt.

Sensible Statistik mit Ausreißereliminierung (nach Grubbs, ISO 5725 und Bremser)

Der χ^2 -Anpassungstests wird in diesem Fall nicht berücksichtigt. Außerdem kann eine Normalverteilung des Datensatzes angenommen werden da die Auswertung im logarithmierten Zahlen-bereich stattfindet (Log10-Normalverteilung).

Ist der Anteil an Bremser-Ausreißer aufgrund sehr guter Materialhomogenität für Ref.meth. $\geq 25\%$ so kann der Bremser-Ausreißer als statistisches Werkzeug nicht angewendet werden. In diesem Fall muss der Datensatz dennoch **auf Normalverteilung geprüft** werden und der **χ^2 -Anpassungstest** wird gemäß den Vorgaben für „chemische Ringversuche“ betrachtet und angewendet. In diesem Fall erfolgt zusätzlich auch die Anwendung der beiden folgenden statistischen Modell gemäß den Vorgaben für „chemische Ringversuche“

- **Sensible Statistik**
- **Robuste Statistik**

Die Auswertung der Daten kann methodenabhängig und/oder methodenunabhängig erfolgen, abhängig von der Art des jeweiligen Ringversuchs.

In Einzelfällen kann der quantitative Zielwert für einen einzelnen quantitativen Parameter im Ringversuch planmäßig < 10 KbE sein (z.B. Probe wurde für den Parameter planmäßig nicht mit Mikroorganismus dotiert). In diesem Fall ist zwar keine statistische Auswertung möglich, dennoch erfolgt eine positive „qualitative Leistungsbewertung“ für alle Labore, die den Zielwert für den betroffenen Parameter richtig bestimmt haben.

Only the sensitive statistics (with outlier elimination) is applied when the application of the Bremser-outlier a statistical tool and the associated sample structure is used.

Sensible statistics with outlier elimination (acc. to Grubbs, ISO 5725 and Bremser)

The χ^2 goodness-of-fit test is not carried out in this case. A normal distribution of the data set can be assumed, since the evaluation takes place in the logarithmic number range (Log10 normal distribution).

If the percentage of “Bremser outliers” is $\geq 25\%$ due to very good material homogeneity for Ref.meth. the “Bremser outlier” cannot be used as statistical tool. In this case, the data set must still be **checked for normal distribution** and the **χ^2 goodness-of-fit test** is considered and applied according to the specification for "chemical PT schemes". In this case, the following two statistical models are also applied according to specifications for "chemical PT schemes"

- **sensible statistics**
- **robust statistics**

The evaluation of the data can be method-dependent and/or method-independent, depending on the type of the respective proficiency test.

In individual cases, the quantitative target value for a single quantitative parameter in the proficiency testing could be < 10 CFU as planned (e.g. sample was not doped with microorganism as planned for the parameter). In this case, no statistical evaluation is possible but a positive "qualitative performance evaluation" is carried out for all laboratories that have correctly determined the target value for the parameter concerned

Verwendete Statistik in der Ringversuchsauswertung (MIKROBIOLOGIE-QUALITATIV)

Applied statistics for evaluation of proficiency testing schemes (MICROBIOLOGY-QUALITATIVE)

Die Vorgehensweise eines qualitativen Ringversuchs in der Mikrobiologie ist analog zur chemischen Aus-

The evaluation of a qualitative proficiency test in the field of microbiology is analogous to the chemical

wertestrategie für qualitative Ringversuche.

Ringversuche Identifizierung (z.B. Mikroorganismus):

Die Vorgehensweise ist analog zur chemischen Auswertestrategie für Ringversuche zur Identifikation.

evaluation strategy for qualitative proficiency tests.

Proficiency testing identification (e.g. micro-organism):

The evaluation of an identification proficiency test in the field of microbiology is analogous to the chemical evaluation strategy for identification proficiency tests.

2.3 Verwendete Statistik in der Ringversuchsauswertung (SENSORIK)

2.3 Applied statistics for evaluation of proficiency testing schemes (SENSORY)

Binomialverteilung

Binominal distribution

$$P_{n,p}(y) = \binom{n}{y} p^y q^{n-y}$$

2.3.1 Rangordnungsprüfung

Die Auswertung des Ringversuchs erfolgt über die Fehlersumme und einer daraus resultierende Fehlernote, die sich über den Mittelwert und der Unsicherheit des Materials berechnen lässt. Zusätzlich erfolgt eine Panelauswertung über den Pagetest (nach DIN 10963).

2.3.1 ranking test

The evaluation of this proficiency testing is based on the error sum and the resulting mark, which can be calculated by the mean and the uncertainty of the material. A panel evaluation takes place via the page-test in addition (according to DIN 10963).

2.3.2 Dreiecksprüfung

Die Auswertung des Ringversuchs erfolgt über die Signifikanzniveau (nach DIN 4120), die sich anhand der Wahrscheinlichkeitsverteilung der Binomialverteilung errechnen lassen. In der Ringversuchsauswertung werden die beiden Signifikanzniveaus α 0,05 sowie α 0,01 dargestellt.

2.3.2 triangle test

The evaluation of this proficiency testing is based on the levels of significance (according to DIN 4120) which can be calculated with the probability distribution of the binominal distribution. In the evaluation of the proficiency testing the levels of significance α 0,05 and α 0,01 are displayed.

2.3.3 Wasserbeschaffenheit DIN EN 1622 (Schwellenwertprüfung):

Teilnehmende Sensorikpanels untersuchen im Ringversuch eine Trinkwasserprobe, bei der ein Fehleroma simuliert wird. Aufgabe ist es den Geruchsschwellenwert (TON) und den Geschmacksschwellenwert (TFN) des Panels zu bestimmen. Das Prinzip der Untersuchung ist eine stufenweise Verdünnung der Probe und der Vergleich mit einem vom DRRR bereitgestellten Referenzwasser. Der Vergleich geschieht mittels gezwungener Dreiecksprüfung nach DIN EN ISO 4120. Um die Schwellenwerte genauer zu ermitteln, werden je Verdünnungsstufe drei Dreiecksprüfungen verkostet. Dadurch kann zu 95 % ausgeschlossen werden, dass ein Ergebnis zufällig entstanden ist. Der individuelle Schwellenwert ist derjenige Verdünnungsfaktor der Verdünnung mit der niedrigsten Deskriptorkonzentration, bei der ein Prüfer gerade noch einen Unterschied zum Referenzwasser bei allen drei Triaden feststellen kann. Ausgewertet werden die logarithmierten Ergebnisse mit den statistischen Werkzeugen der Normalverteilung, da es sich bei den individuellen Prüfer- und Panelschwellenwer-

2.3.3 Water quality acc. to DIN EN 1622 (threshold value test):

The drinking water samples provided for this proficiency testing are artificially flavoured to resemble an ordinary flavour taint. The task for the panels is to determine the panel threshold odour number and threshold flavour number. The examination principle follows a stepwise dilution of the sample and comparing each dilution with the reference water. The method of comparison is the forced triangle test acc to DIN EN ISO 4120. To obtain more statistical sound data the assessors are requested to perform three triangle tests for each dilution. Judging three triads correctly demonstrates the sensory competence within an uncertainty interval of 95 %. The threshold number presents the dilution factor of the dilution with the lowest gustatory or odorous flavor ingredients concentration, where the assessor can determine a difference to the reference water within the uncertainty interval of 95 %. The results will be evaluated with the statistical tools of the normal distribution after calculating the logarithm of the results, because the individual assessor and panel threshold numbers acc. to EN 1622 are logarithmic

ten nach DIN EN 1622 um logarithmisch normalverteilte Daten handelt. Als Zielwert für die Geruchs- und Geschmacksschwellenwerte wird ein Konsenswert über den Ausreißerbereinigten Datensatz (Ausreißertest nach Grubbs) gebildet. Die Bewertung der Panels erfolgt, in Anlehnung an die DIN EN ISO/IEC 17043:2010, mit dem z-score.

2.3.4 Einfach beschreibende Prüfung:

Die genannten Merkmalseigenschaften der Teilnehmer werden je abgefragtem Merkmal gelistet. Inhaltlich gleiche Merkmalseigenschaften werden mittels Gruppierungsbegriffen zusammengefasst. Die Häufigkeit der einzelnen Gruppierungsbegriffe je Merkmal wird ermittelt. Die Gruppierungsbegriffe werden mit ihrer Häufigkeit (Anzahl) gewichtet. Je häufiger eine Merkmalseigenschaft im Ringversuch genannt wurde, desto höher ist die Gewichtung der Merkmalseigenschaft. Die Häufigkeiten dienen als Punkte, die ein Teilnehmer für seine genannten Merkmalseigenschaften je Merkmal bekommt. Die erreichten Punkte werden in Relation zu den maximal erreichbaren Punkten gesetzt. Die maximal erreichbaren Punkte je Merkmal bestehen aus der Summe der fünf am häufigsten genannten Gruppierungsbegriffe, da jeder Teilnehmer fünf Merkmalseigenschaften je Merkmal nennen konnte.

Die Leistung eines Panels wird anhand der erreichten Punkte in Bezug auf die maximal möglichen Punkte je Merkmal beurteilt. Die Kriterien hierfür sind:

- Wenn $\geq 95\%$ der maximal erreichbaren Punkte erreicht wurden, erhalten Sie die Bewertung: „...**mit großem Erfolg teilgenommen**“.
- Wenn $\geq 75\%$ und $< 95\%$ der maximal erreichbaren Punkte erreicht wurden, erhalten Sie die Bewertung: „...**mit Erfolg teilgenommen**“.
- Wenn $< 75\%$ der maximal erreichbaren Punkte erreicht wurden, erhalten Sie die Bewertung: „...**teilgenommen**“.

2.3.5 Sensorik von Packstoffen und Packmitteln (nach DIN 10955) sowie Sensorik von Papier und Pappe (nach DIN EN 1230)

Ziel des Ringversuchs ist die Bestimmung der sensorischen Abweichung eines Prüflebensmittels / Verpackung

normal distributed. The assigned value for the TON and TFN represents the consensus value from the participating panels without outlier. According to the DIN EN ISO/IEC 17043:2010 the panels are evaluated by means of the z-score.

2.3.4 Simple descriptive testing:

The named sensory attributes by the participants are listed for each parameter. Sensory attributes with the same content are pooled with group terms. Then the frequency of the different group terms are determined. The group terms are weighted with their frequency. The more frequently a sensory attribute was named in this proficiency testing the higher is the weighting of the sensory attribute. The weighting is equal to a score the panel achieves for their named sensory attribute for each parameter. The archived score for each parameter is put in relation to the maximum achievable score. The maximum achievable score for each parameter is the sum of the five most frequently named sensory attributes, because every panel could name five sensory attributes for each parameter.

The performance of a panel is evaluated according to the achieved score in relation to the maximum score for each parameter. The criteria is:

- If $\geq 95\%$ of the maximum score is achieved you will get the valuation: „...**has participated with great success**“
- If $\geq 75\%$ and $< 95\%$ of the maximum score is achieved you will get the valuation: „...**has participated with success**“
- If $< 75\%$ of the maximum score is achieved you will get the valuation: „...**has participated**“

2.3.5 Sensory of packaging (acc. DIN 10955) and sensory of paper and board (acc. EN 1230)

The goal of this PT is the determination of the organoleptic deviation of the food simulating matrix / packaging toward a reference sample or calibration substances according to

bzw. Packstoff im Vergleich zu einer Referenzprobe bzw. Kalibriersubstanzen mittels einer 4-stufigen Intensitätsskala nach DIN 10955 bzw. DIN EN 1230, wobei Intensität $I = 0$ (keine wahrnehmbare Geruchs- bzw. Geschmacksabweichung) bis $I=4$ (starke Geruchs-bzw. Geschmacksabweichung).

Die Leistung eines Panels im Ringversuch wird beurteilt, indem die ermittelten Panelmediane mit dem Konsenswert (Median der einzelnen Panelmediane) verglichen werden. Anhand der absoluten Differenz vom Panelmedian zum Konsenswert des entsprechenden Bedarfsgegenstands erfolgt die Bewertung des Panels:

Die Leistung eines Panels wird mit sehr gut bewertet, wenn es innerhalb des Unsicherheitsintervalls (99 %) liegt. Das heißt die absolute Differenz des Panelmedians vom besten Schätzwert liegt innerhalb dieses Unsicherheitsintervalls. Die Leistung eines Panels wird mit gut bewertet wenn die absolute Differenz des Panelmedians vom besten Schätzwert kleiner oder gleich 1,5 ist. Dies ist darin begründet, dass nach DIN 10955 bzw. DIN EN 1230 dann übereinstimmende Ergebnisse vorliegen, wenn die einzelnen Ergebnisse nicht mehr als 1,5 vom Median abweichen. Dieses Kriterium wird für die Probe die mit Kalibriersubstanzen geliefert wird auf 0,5 verschärft, da Kalibriersubstanzen mitgeliefert werden und somit eine deutlich bessere Übereinstimmung erwartet wird.

Die Auswertung und Leistungsbewertung in diesem Ringversuch wird nur für Panels mit mindestens drei Prüfern vorgenommen, für die maximale Aussagekraft sollten die Labore aber mit mindestens 6 Prüfpersonen teilnehmen.

Die Abweichung des Panelmedians zum Konsenswert wird in der tabellarischen Übersicht aller Panelergebnisse zahlenmäßig dargestellt.

Bewertet wird die Leistung der Panels jeder der drei Verkostungen im Ringversuch. Aufbauend auf diese Leistungskriterien werden für den Ringversuch folgende Bewertungskriterien für die Gesamtbewertung festgelegt:

- Hat das Panel in allen Verkostungen eine sehr gute Leistung gezeigt, erhält das Panel die Bewertung: „... mit großem Erfolg teilgenommen“.
- Hat das Panel in allen Verkostungen eine gute bis sehr gute Leistung gezeigt, erhält das Panel die Bewertung: „... mit Erfolg teilgenommen“.
- Hat das Panel nicht in allen Verkostungen eine gute bis sehr gute Leistung gezeigt, erhält das Panel die Bewertung: „... teilgenommen“.

Bewertet wird die Gesamtleistung eines Panels hinsichtlich der Bewertung der sensorischen Abweichung einer Probe im Vergleich zu einer Referenz bzw. im Vergleich zu den Kalibriersubstanzen. Hat ein Panel nicht alle drei Verkostungen durchgeführt, ist keine Gesamtbewertung möglich. Die Leistung der einzelnen Prüfpersonen wird nicht bewertet. Die einzelnen Prüfpersonen erhalten ein Teilnehmerzertifikat.

the scale (acc. to DIN 10955 or DIN EN 1230) from $I=0$ (no taste / odour deviation) to $I=4$ (strong taste / odour deviation).

The performance of a panel in this PT is evaluated while the determined panel medians are compared to the consensus value (median of the single panel medians).

The performance rating for the single panel is done on the basis of the absolute difference between the panel median and the consensus value for the sample. The performance of a panel is rated as very successful if the absolute deviation of the panel median to the assigned value is smaller or equal the uncertainty interval (99 %). The performance of a panel is rated as successful if the absolute deviation of the panel median to the assigned value is smaller or equal 1,5. This is due to the fact that the standard DIN 10955 and EN 1230 states that consistent results are produced if the single assessors results do not deviate more than 1.5 from the median.

This criterion is tightened for the evaluation of the sample which is delivered with calibration substances to 0.5, because an improved concordance can be expected due to the provided calibration substances.

Only for panels with at least 3 assessors a performance evaluation is calculated in this proficiency testing. For the maximum informative value a laboratory should participate with at least 6 assessors. The deviation of the panel median to the consensus value is numerically described and tabulated.

The performance of the panels is evaluated for each of the three testing sessions in this PT. Based on this performance criteria the overall performance evaluation of the panel is determined:

- *If the panel showed a very successful performance in all testing sessions the panel will get the valuation: "...has participated with great success".*
- *If the panel showed a successful to a very successful performance in all testing sessions the panel will get the valuation: "...has participated with success".*
- *If the panel did not show a successful to a very successful performance in all testing sessions the panel will get the valuation: "...has participated".*

The overall performance of a panel is evaluated concerning their assessment of the organoleptic deviation of the sample compared to a reference or the calibration substances. If a panel did not carry out all three testings an overall performance evaluation is not possible. The performance of the single assessors is not evaluated. The single assessor receive a certificate of participation.

2.4 Verwendete Statistik in der Ringversuchsauswertung (SMALL ILC)

2.4 Applied statistics for evaluation of proficiency testing schemes (SMALL ILC)

Liegen für einen Ringversuch weniger als 7 Anmeldungen (z.B. 5 oder 6) vor handelt es sich um einen Small ILC.

Die Standardvorgehensweise zur Auswertung (chemischer Ringversuche) wird auch für Small ILC's immer angewendet. Für jeden Small ILC erfolgt eine gesonderte Plausibilitätsprüfung der Auswertung, die nach folgender Vorgehensweise abläuft:

Ergibt sich für die Standardvorgehensweise des DRRR und einen externen Referenzwert (z.B. Expertenlabor, Einwaageprotokoll, etc.) die gleiche Laborbewertung (Einstufung: zufriedenstellend, fragwürdig und unbefriedigend), wird die DRRR Standardvorgehensweise genutzt, da kein Unterschied zum externen Referenzwert vorliegt und eine Berechnung anhand der Ringversuchsdaten möglich ist.

Ergibt sich für die Standardvorgehensweise des DRRR und den externen Referenzwert nicht die gleiche Laborbewertung (Einstufung: zufriedenstellend, fragwürdig und unbefriedigend), kann die Auswahl der Statistik angepasst (z.B. mit einem höheren χ^2 -Wert) oder ein externer Referenzwert zur Definition von Mittelwertlage und Streumaß für den Ringversuch herangezogen werden.

If there are less than 7 registrations (i.e. 5 or 6) for a proficiency testing, it is a Small ILC.

The standard procedure for evaluation (chemical proficiency testing) is always used for small ILC's. A separate plausibility check of the evaluation is carried out for each Small ILC, which is carried out according to the following procedure:

If the standard DRRR procedure and the external reference value (e.g. expert laboratory, weighting protocol and so on) share the same laboratory assessment (classification: satisfactory, questionable and unsatisfactory), the DRRR standard procedure shall be used as there is no difference from the external reference value and calculation based on the proficiency test data is possible.

If the standard DRRR procedure and the external reference value do not result in the same laboratory evaluation (classification: satisfactory, questionable and unsatisfactory), the selected statistic can be changed (e.g. with a higher χ^2 -value) or an external reference value is used for the definition of the best estimate and the standard deviation of the proficiency test.

3. Homogenität und Stabilität

3. Homogeneity and stability

3.1 Homogenitätstest nach BCR/48/93

Stichproben-Homogenität H_S

Stichproben-Homogenität H_S ist das Intervall eines erwarteten Messbereichs von 10 repräsentativen Stichproben ausgedrückt als erweiterte Unsicherheit (95,5%) U_{ex} von jeweils einer Einzeluntersuchung

$$H_S = S_{(\text{Stichprobe})} \cdot t$$

Analytische Homogenität H_a

Analytische Homogenität H_a der Einzelprobe ist das Intervall eines erwarteten Messbereichs ausgedrückt als erweiterte Unsicherheit (95,5%) U_{ex} von 10 Einzeluntersuchungen an 1 repräsentativen Stichprobe. Da angenommen werden kann, dass eine Einzelprobe ideal homogen ist, resultiert die tatsächlich gemessene erweiterte Unsicherheit (95,5%) U_{ex} ausschließlich aus der analytischen Messunsicherheit.

$$H_a = S_{(\text{analyt})} \cdot t$$

Material-Homogenität H_M

Material-Homogenität H_M ist das resultierende erweiterte Unsicherheitsintervall (95,5%) U_{ex} aus der Differenz der Wurzel der quadrierten Standardabweichungen s ermittelt aus der Bestimmung der Stichproben-Homogenität H_S und der Analytischen Homogenität H_a .

Die Standardabweichung s_{Material} berechnet sich nach unten stehender Gleichung:

$$H_M = S_{\text{Material}} \cdot t$$

$$S_{\text{Material}} = \sqrt{S_{\text{Stichprobe}}^2 - S_{\text{analyt}}^2}$$

3.1 Homogeneity test according BCR/48/93

Homogeneity of random sample batch H_S

This total uncertainty H_S expresses the extended uncertainty (95,5%) U_{ex} of 10 measurements out of 10 representative random samples (in between homogeneity).

Analytical homogeneity H_a

The analytical uncertainty H_a expresses the extended uncertainty (95,5%) U_{ex} of 10 measurements out of 1 representative random sample. It is assumed, that one sample itself shows no inhomogeneity.

Material homogeneity H_M

The material homogeneity H_M expresses the extended uncertainty (95,5%) U_{ex} . It is computed out of the square root of the squared homogeneity of the random sample batch H_S and analytical homogeneity H_a .

The standard deviation s_{Material} is calculated according the below equation:

Homogenitätstest für sensorische Referenzmaterialien

Die Homogenität zwischen den Probenpackungen erfolgt über ein spezielles Verfahren in Anlehnung

homogeneity for sensory reference materials

The checking of homogeneity between sample packages of sensory samples results from a special

an das Verfahren von BCR/48/93. Es wird die Unterscheidbarkeit der einzelnen Konzentrationsstufen geprüft und ob bei Wiederholung der sensorischen Tests für die einzelnen Konzentrationsstufen reproduzierbare Ergebnisse erzielt werden. Für den Homogenitätstest werden in der Regel 72 sensorische Analysen durchgeführt. Über das statistische Verfahren werden Wiederholbarkeit, Vergleichbarkeit als auch die Unsicherheit bestimmt. Die Unsicherheit einer Konzentrationsstufe ist das Maß der Variabilität von 12 Messungen um den Mittelwert.

procedure according to the method BCR/48/93. In this process was checked, if the samples are discriminated or not and if reproducible results for each sample by repeating the sensory testing are obtained. In general 72 sensory analyses are carried out for the homogeneity testing. The repeatability, the comparability as well as the uncertainty are identified by this statistical procedure. The uncertainty of each sample presents the variability of twelve measurements around the mean value.

3.2 Homogenitätstest nach ISO 13528

Für den Homogenitätstest werden je nach Batchgröße an 5,7 oder 10 repräsentativen Stichproben Doppelbestimmungen durchgeführt. Anhand dessen werden s_x , s_w und s_s des Materials berechnet.

Für chemische Prüfgegenstände gilt: Wenn die between-sample standard deviation s_s kleiner ist als 30 % einer errechneten Prüfgröße werden die Proben als homogen eingestuft. Die Prüfgröße errechnet sich über alle bisherigen Ringversuche eines Materialtyps indem die relative mittlere Standardabweichung sowie U_{ex} der relativen mittleren Standardabweichung über alle Ringversuche dieses Materialtyps ermittelt werden. Diese Relativgrößen werden dann bezogen auf die Ist-Werte des Materials in Absolutwerte umgerechnet und addiert. Wenn diese Regel nicht eingehalten werden kann, wird geprüft, ob im Einzelfall die erhöhte Streuung für den Materialtyp üblich ist.

Für mikrobiologische Prüfgegenstände gilt: Die erweiterte Unsicherheit des Probenmaterials ist $< 0,5 \log$ (Homogenitätskriterium).

Definition der Variablen (ISO 13528)

t = Proben ($t=1,2,\dots,g$)

k = Untersuchungen einer Doppelbest. ($k=1,2$)

$X_{t..}$ = Mittelwert der Doppelbest. ($x_{t..} = (x_{t,1} + x_{t,2}) / 2$)

w_t = Differenz zweier Werte der Doppelbest. ($w_t = |x_{t,1} - x_{t,2}|$)

$\bar{x}_{...}$ = Mittelwert aller Doppelbest. ($\bar{x}_{...} = \sum \bar{x}_{t..} / g$)

Standardabweichung der Probenmittelwerte s_x

3.2 Homogeneity test according ISO 13528

The homogeneity test is carried out at 5, 7 or 10 representative random samples in double determination depending on the amount of samples in the batch. According to this s_x , s_w and s_s for the material are calculated.

For chemical testing samples the following rule applies: If the between-sample standard deviation s_s is smaller than 30 % of a test statistic the samples are classified as homogenous. The test statistic is calculated of all previous proficiency testings with this material type determining the relative mean standard deviation as well as U_{ex} of the relative mean standard deviation of all previous proficiency testings with this material type. The relative values are referred to absolute values related to the actual value of the material and are added to obtain the test statistic. When this rule cannot be followed it will be proved if the higher variance is usual for this material type.

For microbiological testing samples the following rule applies: The extended uncertainty of the material is below $0,5 \log_{10}$ (criteria for homogeneity).

Definition of the variables (ISO 13528)

t represents the samples ($t=1,2,\dots,g$)

k represents the test portion ($k=1,2$)

$X_{t..}$ defines the sample averages ($x_{t..} = (x_{t,1} + x_{t,2}) / 2$)

w_t defines between-test-portion ranges ($w_t = |x_{t,1} - x_{t,2}|$)

$\bar{x}_{...}$ defines the general average ($\bar{x}_{...} = \sum \bar{x}_{t..} / g$)

Standard deviation of sample averages s_x

$$s_{Stichprobe} = s_x = \sqrt{\sum (x_{t..} - \bar{x}_{...})^2 / (g - 1)}$$

Within-sample standard deviation s_w

Within-sample standard deviation s_w

$$s_{analyt} = s_w = \sqrt{\sum w_t^2 / (2g)}$$

Between-sample standard deviation s_s

Between-sample standard deviation s_s

$$s_{Material} = s_s = \sqrt{s_x^2 - (s_w^2 / 2)}$$

3.3. Homogenitätstest über Einwaageprotokoll

Die Homogenitätsprüfung erfolgt mittels Gravimetrie anhand der Einwaage. Zur Bestimmung der Materialstandardabweichung $s(\text{Material})$ wird das Einwaageprotokoll verwendet, da die Proben einzeln dotiert wurden. Die Gesamtstandardabweichung $s(\text{Material})$ wird relativ ermittelt anhand der Gesamteinwaage.

Wenn die Standardabweichung ($s(\text{Material})$) der Material-Homogenität (HM) kleiner ist als 30 % der Standardabweichung für die Eignungsbewertung (SDPA) werden die Proben als homogen eingestuft.

3.3. Homogeneity test acc. weigh-in protocol

The homogeneity test is carried out by gravimetric procedure according to the weigh-in protocol, since each sample was spiked individually. The total standard deviation $s(\text{Material})$ according to the gravimetric procedure is calculated in relative terms according to the total quantity weighed-in.

If the standard deviation ($s(\text{material})$) of the material homogeneity (HM) is smaller than 30 % of the standard deviation for proficiency assessment (SDPA) the samples are classified as homogenous.

3.4. Post-Hoc-Homogenitätstest

Wird nach der zweifaktoriellen Varianzanalyse mit Messwiederholung auf dem Signifikanzniveau α 0,05 gemäß ISO GUIDE 35 durchgeführt.

3.4. Post-Hoc-Homogeneity test

The Post-Hoc-homogeneity test according ISO GUIDE 35 is carried out with the two-way ANOVA with replicates on the level of significance α 0,05.

3.5. Stabilitätsprüfung mittels Referenzanalytik

Für den Stabilitätstest wird mittels Referenzanalytik wird an zwei Proben jeweils eine Einzelbestimmung durchgeführt. Es folgen zwei Prüfschritte:

Prüfung der Präzision: die Differenz zwischen den beiden Werten der Einzelbestimmung soll dabei kleiner sein als das Kriterium für die Präzision. Das Kriterium für die Präzision ist r . Falls r nicht vorhanden ist, wird stattdessen die gewichtet-kombinierte Standardabweichung aus der Zusammenfassung des letzten Ringversuchs / der letzten Ringversuche verwendet.

Prüfung der Richtigkeit: Prüfung ob der Mittelwert der beiden Einzelbestimmungen innerhalb der Warn- bzw. Eingriffsgrenze liegt. Dabei errechnen sich diese beiden Grenzen folgendermaßen:

Die obere bzw. untere Warngrenze errechnet sich aus dem aktuellen Referenzwert für das Material+ bzw. $-r$ (falls r nicht vorhanden wird auch in diesem Fall die gewichtet-kombinierte Standardabweichung aus der Zusammenfassung des letzten Ringversuchs / der letzten Ringversuche verwendet).

Die obere bzw. untere Eingriffsgrenze errechnet sich aus dem aktuellen Referenzwert für das Material + bzw. $-R$ (falls R nicht vorhanden wird die erweiterte Unsicherheit U_{ex} des Materials aus der Zusammenfassung des letzten Ringversuchs / der letzten Ringversuche verwendet).

3.5. Stability test with reference analysis

For the stability test, an individual determination is carried out on two samples by means of reference analysis. This is followed by two test steps:

Precision check: the difference between the two values of the individual determination should be smaller than the criterion for precision. The criterion for precision is r . If r is not available, the weighted combined standard deviation from the summary of the last proficiency test(s) is used instead.

Verification of correctness: Verification whether the mean value of the two individual determinations lies within the warning or intervention limit. These two limits are calculated as follows:

The upper or lower warning limit is calculated from the current reference value for the material+ or $-r$ (if r is not available, the weighted combined standard deviation from the summary of the last proficiency test / the last proficiency tests is also used in this case).

The upper or lower intervention limit is calculated from the current reference value for the material + or $-R$ (if R is not available, the extended uncertainty U_{ex} of the material from the summary of the last proficiency test/ the last proficiency tests is used).

3.6. Stabilitätstest nach ISO 13528

Für den Stabilitätstest wird an zwei Proben mindestens 1 Doppelbestimmung durchgeführt. Es folgen gemäß ISO 13528 Anhang B folgende Prüfschritte:

Der Mittelwert der Homogenitätsprüfung wird mit dem Mittelwert der Stabilitätsprüfung verglichen.

Die Proben können als ausreichend stabil angesehen werden wenn folgendes Kriterium erfüllt ist:

$$\left| \bar{y}_1 - \bar{y}_2 \right| \leq 0,3\sigma_{pt} + 2\sqrt{u^2(\bar{y}_1) + u^2(\bar{y}_2)}$$

Die Streuung (between standard deviation s_s) der Stabilitätsprüfung sollte folgendes Kriterium erfüllen:

$$s_s \leq 0,3 \sigma_{pt}$$

3.6 Stability test according ISO 13528

For the stability test at least 1 double determination is carried out on two samples. According to ISO 13528 Annex B, the following test steps follow:

Compare the general average of the measurements obtained in the homogeneity check with the general average of the results obtained in the stability check. The samples may be considered to be adequately stable if:

The variance (between standard deviation s_s) of the stability test shall meet the following criterion:

4. Messung der Laborleistung

3. Measurement of performance

z-score [4]

$z \leq 2,00 $	zufriedenstellende Laborleistung
$ 2,00 < z < 3,00 $	fragwürdige Laborleistung
$z \geq 3,00 $	unbefriedigende Laborleistung

Der z-score wird auf zwei Nachkommastellen gerundet.

$$z - score = \frac{m - m_{best}}{s_R}$$

z-score [4]

$z \leq 2,00 $	performance is satisfactory
$ 2,00 < z < 3,00 $	performance is questionable
$z \geq 3,00 $	performance is unsatisfactory

The z-score is rounded to two decimal places.

z'-score [4]

$z' \leq 2,00 $	zufriedenstellende Laborleistung
$ 2,00 < z' < 3,00 $	fragwürdige Laborleistung
$z' \geq 3,00 $	unbefriedigende Laborleistung

Der z'-score wird auf zwei Nachkommastellen gerundet.

$$z' - score = \frac{m - m_{best}}{\sqrt{s_{best}^2 + s_{Material}^2}}$$

z'-score [4]

$z' \leq 2,00 $	performance is satisfactory
$ 2,00 < z' < 3,00 $	performance is questionable
$z' \geq 3,00 $	performance is unsatisfactory

The z'-score is rounded to two decimal places.

CRD-Wert

Die Berechnung der kritischen Differenz (*CD*) und des CRD-Wertes einer Untersuchungsmethode dient dem Vergleich des Labormittelwertes *m* mit dem Besten Schätzwert für den wahren Wert *m_{best}*, wobei ausschließlich auf den Unterschied zwischen den Laboratorien abgehoben wird. Der Beitrag der jeweiligen Laborstreuung zu *R* (Vergleichbarkeit) interessiert an dieser Stelle nicht. Da *R* bereits *r* (Wiederholbarkeit) enthält, wird der Betrag von *r* aus *R* (in Abhängigkeit von *n*, der Freiheitsgraden) abgezogen. Da es sich hierbei um die Überschreitung eines Höchstwertes bzw. die Unterschreitung eines Mindestwertes handeln kann, muss eine zweiseitige Fragestellung behandelt werden.

CRD-Wert $\leq 1,00$ = Laborwert ist zufriedenstellend

CRD-Wert $> 1,00$ = Laborwert ist unbefriedigend

CRD-value

The calculation of the critical difference (*CD*) and the CRD-value depending on the used testing methods is used for the comparison of the lab mean value *m* and the best estimate of the true value *m_{best}*. These values only focuses on the comparison of *R* (reproducibility) exempt from *r* (repeatability) of the labs. Therefore *R* gets reduced by *r* in dependence of *n*. It has to be discussed a double sided question because it is possible to get a overshooting of the maximum value or a undershooting of the minimum value.

CRD-value $\leq 1,00$ = laboratory value is satisfactory

CRD-value $> 1,00$ = laboratory value is unsatisfactory

Der CRD-Wert wird auf zwei Nachkommastellen gerundet.

The CRD-value is rounded to two decimal places.

Die Gleichung für diese zweiseitige Fragestellung nach § 64 LFGB ist nachstehend aufgeführt.

The equation of the double sided question according to § 64 LFGB is as follows:

$$CD = \frac{1}{\sqrt{2}} \cdot \sqrt{(2,8s_R)^2 - (2,8s_r)^2 \left(\frac{n-1}{n}\right)}$$

$$CD = \frac{1}{\sqrt{2}} \cdot \sqrt{R^2 - r^2 \left(\frac{n-1}{n}\right)}$$

Für eine Doppelbestimmung gilt außerdem $n = 2$:

Because of the double determination $n = 2$:

$$CD = \frac{1}{\sqrt{2}} \cdot \sqrt{R^2 - \frac{r^2}{2}}$$

$$CD = 0,707 \sqrt{R^2 - \frac{r^2}{2}}$$

$$CRD - \text{Wert} = \frac{m - m_{\text{best}}}{CD (\text{Methode})}$$

**Anwendung der
Standardabweichung zur Eignungsbewertung
(„SDPA min“ / „SDPA“)**

Im Rahmen der Ringversuchsentwicklung wird für jeden Ringversuch und Parameter eine minimale Standardabweichung zur Eignungsbewertung („SDPA min“) und eine maximale Standardabweichung zur Eignungsbeurteilung („SDPA“) festgelegt.

Die im Ringversuch ermittelte Standardabweichung (s_{best}) darf nur dann für die Laborbewertung herangezogen werden, wenn

- s_{best} die „SDPA min“ nicht unterschreitet
- s_{best} die „SDPA“ nicht überschreitet.

**Usage of the
Standard deviation for Proficiency assessment
(“SDPA min” / “SDPA”)**

During the planning process for each proficiency test and parameter, a minimum standard deviation for Proficiency assessment (“SDPA min”) and a maximum standard deviation for Proficiency assessment (“SDPA”) is defined.

The calculated standard deviation (s_{best}) of the proficiency test can only be used for the laboratory performance evaluation, if

- s_{best} does not fall below “SDPA min”
- s_{best} does not exceed “SDPA”

Wird „SDPA min“ bzw. „SDPA“ von der berechneten Standardabweichung "sbest" unter bzw. überschritten wird "sbest" durch "SDPA min" bzw. "SDPA" ersetzt, ansonsten ist "sbest" als Streumaß für die Bewertung über den z'-score zu verwenden.

If the calculated standard deviation "sbest" falls below or exceeds this limit, "sbest" is replaced by "SDPA min" resp. "SDPA", otherwise "sbest" is used as estimator for the z'-score evaluation in the Proficiency test.

5. Umgang mit Kundendaten

4. Handling of customer data

Umgang der DRRR GmbH mit übermittelten Ergebnissen und Kundendaten im Ringversuch

- 1) Generell werden Ergebnisse / Kundendaten durch die DRRR GmbH nicht geändert, sondern so in der Auswertung berücksichtigt, wie diese durch den Kunden selbst übermittelt wurden.
- 2) Es erfolgt keine Informationsänderung bzw. Informationsergänzung von Kundeninformationen seitens der DRRR GmbH.
- 3) Es werden lediglich redaktionell notwendige, formale Änderungen vorgenommen, sowie Informationen entfernt, die den Datenschutz des Kunden betreffen.
- 4) Ebenso behält sich die DRRR GmbH vor, unsachliche Bemerkungen des Kunden zu entfernen, die nicht Ergebnis- oder Ringversuchsbezogen sind.

Handling of submitted results and further customer data by DRRR during a PT scheme

- 1) Generally, submitted results and further customer data won't be changed by DRRR GmbH but will be used during the evaluation as it was submitted by the client.
- 2) There are no changes or modifications of customer information on the part of DRRR GmbH.
- 3) There are only editorial modifications or removal of client data for reasons of data protection.
- 4) Furthermore DRRR GmbH reserves the right to remove unobjective personal remarks of customers not referring to the results or to the corresponding PT scheme.

Verzeichnis der verwendeten Standards / Normen / Register of used standards / norms

1. DIN EN ISO /IEC 17025

*Allgemeine Anforderungen an die Kompetenz
von Prüf- und Kalibrierlaboratorien*

2. ILAC-G13

*Guidelines for the Requirements for the Com-
petence of Providers of Proficiency Testing
Schemes*

4. ISO 13528

*Statistical methods for use in proficiency test-
ing by interlaboratory comparison*

5. DIN 38402-45

*Ringversuche zur externen Qualitätskontrolle
von Laboratorien*

6. DIN 5725-2

*Genauigkeit und (Richtigkeit und Präzision)
von Messverfahren und Messergebnissen,
Teil 2*

7. DIN 5725-4

*Genauigkeit und (Richtigkeit und Präzision)
von Messverfahren und Messergebnissen,
Teil 4*

8. IDF 135B:

*Precision characteristics of analytical Methods
– Outline of collaborative study procedure*

9. BCR 48/93

*Guidelines for the production and certification
of BCR reference materials*

verwendete Abkürzungen / used abbreviation:

DIN Deutsches Institut für Normung e. V.
EN Europäische Norm

IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ILAC	International Laboratory Accreditation Cooperation
IDF	International Dairy Federation

Anhang zum statistischen Protokoll

Fragestellung	Antwort	Literaturstelle
Wann sind zwei Mittelwerte statistisch unterscheidbar?	Überdecken sich die die Vertrauensbereiche nicht, so besteht zwischen den Parameters ein echter Unterschied: $H_0: \mu_1 = \mu_2$ lässt sich für $n_1 > 10$ und $n_2 > 10$ auf dem 5%-Niveau ablehnen, sobald sich die beiden 95%-Vertrauensbereiche nicht überlappen.	Sachs L. (2002): Angewandte Statistik; 11. Aufl. Springer Verlag, S. 358 http://books.google.de/books?id=MExrugxM1loC&pg=PA358&dq=vertrauensbereiche+nicht+%C3%BCberlappen&cd=4#v=onepage&q=vertrauensbereiche%20nicht%20%C3%BCberlappen&f=false